Applied Machine Learning & Data Science Projects

Data Science 435 / 535 & Computer Science 449 / 549 • Spring 2021

Time & Location:
Mondays & Wednesdays 5:00pm - 6:30pm
Location: Fully Online (Synchronous & Asynchronous)

Instructors:
Genevera Allen (gallen@rice.edu); Office Hour: Wednesdays 4:00-5:00pm
Su Chen (sc131@rice.edu); Office Hour: Friday 12:00-1:00pm
Arko Barman (ab139@rice.edu); Office Hour: TBD

D2K Lab Administrator:
Jennifer Sanders (Jennifer.Sanders@rice.edu or d2k@rice.edu)

Course Description:
In this project-based course, student teams will complete semester-long data science research or analysis projects selected from a variety of disciplines and industries. Students will also learn best practices in data science.

Course Overview:
This is a project course where student teams will complete semester-long data science research or analysis projects selected from a variety of disciplines and industries. Each student team will have a Rice instructor mentor and a sponsor mentor. Student teams are required to meet weekly outside of class with their mentors. During the semester, teams will work towards project milestones and present their progress in interim reports, oral presentations, and software checks. At the end of the semester, each team will produce a final project report with accompanying software and give a presentation on their work at the D2K (Virtual) Showcase. This course will also have a lecture component (online synchronous and asynchronous videos) where students will learn how to design a data science pipeline and best practices in data science with emphasis placed on data scientific reproducibility and reliability.

Course Objectives & Outcomes:
In this course, students will engage in experiential learning where they learn best practices in data science by working on real-world data science projects. Students will learn how to wrangle, explore, and model their data as well as how to design a reproducible and reliable data science pipeline, validate their data science findings, and discuss the ethical implications of their work. Through these projects, students will also learn team-work, leadership, project management, and communication skills. Students completing this course should be able to (i) successfully complete a data science research or analysis project, (ii) successfully devise and employ a reproducible and reliable data analysis pipeline, (iii) work collaboratively on a team to complete a data science project, write reports, develop software, and give presentations, and (iv) effectively communicate their findings to both a general and a technical audience.

DSCI 435 / 535 vs. COMP 449 / 549:
Computer science students should enroll in COMP 449 or COMP 549, and all other students should enroll in DSCI 435 or DSCI 535. Graduate students should enroll in the 500-level course and undergrads in the 400-level course. Graduate students are expected to help lead their team as well as provide major technical contributions to the team project.

Course Application & Project Placement:
This course requires an application. While students can pre-register for the course before the semester begins, registration will only be finalized after students have applied and been placed on a team during the first week of the semester. The course application will ask students to highlight their data science

background and experience as well as list their project preferences and interests. Students will be selected for specific projects based on their background and skill set as well as their project preferences.

Class Time & Attendance:
Class time will be used for lectures, discussions, team check-ins and presentations, software checks, and also mentoring time for the instructors to meet with teams. These are indicated on the course schedule. Students are expected to attend classes and actively participate in all aspects. While attendance will not be taken and strictly enforced, excessive absences will reduce the class participation portion of the grade. Towards the beginning of the semester, there will be a series of lectures including both asynchronous online videos followed by short quizzes as well as in-class synchronous lectures, demonstrations, and discussions. During the later portion of the semester, the class will split into two groups of teams that will meet either during class time on Mondays OR Wednesdays. This portion of the course will entail more focused mentoring during class time. Students are only expected to attend class on their assigned day. All presentations, software checks, and reports will be due by 5pm on the team's assigned day.

Coursework:
The major work of this course will be to conduct a data science research or analysis project. Each team will present their work through presentations at the D2K (Virtual) Showcase and through a final report with accompanying software due at the end of the semester. Throughout the semester, shorter interim reports, presentations, and software checks will be due to give each team feedback and guidance on how best to prepare their final report and presentation. Further details and instructions on the reports, presentations, and software checks as well as grading rubrics are given in separate documents.

Course Expectations:
Students are expected to work at least 10 hours a week on their project outside of class. Each week outside of class, teams are required to have one meeting with their mentors and at least one separate meeting with just team members. Attendance at team meetings is required. Additionally, each team member is expected to contribute to all aspects of the data science project including technical data science contributions, software development, and preparing the written report and presentations.

Grading Overview:
Individual: 30%
- Peer evaluations: 10%
- Individual project contributions: 10%
- Class Participation: 10%

Team: 70%
- Interim Evaluations: 15%
- Final Report: 25%
- Final Presentation: 15%
- Final Software: 15%

Grading rubrics as well as details on all coursework are given in separate documents.

Computing Resources:
The D2K Lab has partnered with Amazon Web Services (AWS) to allot each team $500 of computing credits. Teams can request additional credits as needed. Please use this resource responsibly.

Course Schedule:

Date	Class Time	Assignments	Additional Notes
25-Jan	Projects Intro + DS Pipeline	Course & Project Applications due **26-Jan @ 12pm**	
27-Jan	Meet Teams! + Data Security	Watch Data Security lecture video before class	
1-Feb	DS Pipeline Design	Review Case Study posted on Canvas before class	
3-Feb	Validation, Reproduciblity & Reliablity		
8-Feb	Github & Project Management	Watch Github & Proj Management lecture videos before class	
10-Feb	Software Design & Cloud Computing	Watch Software Design lecture video & complete AWS training before class	
15-Feb	Lightening Team Check-ins (Whole Class)		2-minute presentations introducing project
22-Feb & 24-Feb	**Initial Presentations**	**Initial Report + Catme I due @ 5pm**	Initial Report & Presentation should focus on: project motivation and objectives, background lit review on domain knowledge & potential tehcniques, data description, & data science pipeline design plan
3-Mar	Data Exploration & Interpretable ML	Watch Data Exploration and Interpretable ML lecture videos before class	
8-Mar & 10-Mar	**Software Check I**	(Optional Initial Report Revision due)	Revisions can yield up to a 15% grade increase
15-Mar & 17-Mar	Team Check-ins + Ethics Moment (Privacy)	Watch Intro to DS Ethics video before class	
22-Mar & 24-Mar	Team Check-ins + Ethics Moment (Fairness)		
29-Mar & 31-Mar	**Midterm Presentations**	**Midterm Report + Catme II due @ 5pm**	Midterm Report & Presentation should include aspects from inital report and additionally focus on: data wrangling, data exploration, and any initial modeling results
5-Apr & 7-Apr	Team Check-ins + Ethics Moment (Transparancy, Explainability & Interpretability)		
12-Apr & 14-Apr	**Software Check II**	(Optional Midterm Report Revision due)	Revisions can yield up to a 15% grade increase
19-Apr & 21-Apr	Team Check-ins + Ethics Moment (Robustness & Reliabilty)		
26-Apr & 28-Apr	Draft Final Presentations		Draft 1-min videos required & 5-min videos optional
5-May	**D2K Virtual Showcase (Live!)**	**Final Presentation Videos due 3-May @ 5pm**	
12-May	END OF SEMESTER	**Final Report + Final Software + Catme III due @ 5pm**	
			Gray Fill = Weeks where Monday and Wednesday teams meet separately

Students with a documented disability requiring accommodations should speak with the instructors during the first two weeks of class. Students who reside in time zones that make attendance during classes or team meetings challenging should speak with the instructors during the first two weeks of class.

This syllabus is subject to change with reasonable advance notice by the instructors.